MBMG
Metabarcoding & Metagenomics

Metabarcoding and Metagenomics 10: 377–398 (2026)
DOI: 10.3897/mbmg.10.184566

Research Article

Profiling the eukaryotic diversity over a year in a lake ecosystem through short- and long-read metabarcoding

Marina Chauvet[1], Paul Terzian[1], Anne Moné[1], Didier Debroas[1], Gisèle Bronner[1], Cécile Lepère[1]

1 *CNRS, Laboratoire Microorganismes: Génome Et Environnement, Université Clermont Auvergne, Clermont-Ferrand 63000, France*
Corresponding author: Marina Chauvet (chauvet.marina.pro@gmail.com)

Abstract

Microbial eukaryotes play a vital role in biogeochemical cycles and aquatic food webs. Over past decades, their taxonomic diversity has been investigated using short-read DNA metabarcoding, which, while effective, is hampered by the limited resolution of the targeted regions. In this study, we utilised both short-read (Illumina) and long-read (PacBio) metabarcoding approaches to analyse eukaryotic diversity in a lake ecosystem over one year, allowing for a direct comparison of the results obtained from each method. Our metabarcoding analysis revealed a high degree of congruence between long-read and short-read data at broad taxonomic levels. However, at finer taxonomic scales, such as the genus level, long-read sequencing achieved higher resolution, enabling more precise identification of microbial eukaryotes. This enhanced taxonomic resolution proved especially valuable for tracking seasonal dynamics within key groups, including ciliates and chytrid fungi. The resolution offered by long-read sequencing provides a more detailed picture of the microbial eukaryotic community, thereby facilitating the exploration of ecological interactions at a finer scale. For example, this approach allowed us to monitor the dynamics of the chytrid genus *Zygorhizidium* in relation to various diatom genera, which are primary targets for chytrid parasitism.

Key words: Long-read sequencing, metabarcoding, microbial eukaryotes, short-read sequencing, temporal dynamics

OPEN ACCESS

Academic editor: Thorsten Stoeck
Received: 7 January 2026
Accepted: 7 April 2026
Published: 24 April 2026

Citation: Chauvet M, Terzian P, Moné A, Debroas D, Bronner G, Lepère C (2026) Profiling the eukaryotic diversity over a year in a lake ecosystem through short- and long-read metabarcoding. Metabarcoding and Metagenomics 10: e184566. https://doi.org/10.3897/mbmg.10.184566

Copyright: © Marina Chauvet et al.
This is an open access article distributed under terms of the Creative Commons Attribution License (Attribution 4.0 International – CC BY 4.0).

Introduction

Over the past decades, DNA sequencing has become an indispensable tool for revealing and assessing the diversity and ecology of microbial eukaryotes and has enabled the discovery of organisms across all taxonomic ranks (e.g. Pawlowski et al. (2012); Heger et al. (2018)). In aquatic ecosystems, sequencing has uncovered an unexpected diversity of eukaryotic microorganisms (e.g. Debroas et al. (2015); de Vargas et al. (2015); Massana et al. (2015)). These eukaryotes, ranging from photosynthetic algae to heterotrophic protists, play crucial roles in nutrient cycling, primary production and the regulation of microbial populations through predation or parasitism (e.g. Worden et al. (2015); Sures et al. (2017)). However, despite their diversity, widespread distribution and involvement in biogeochemical cycles, microbial eukaryotes still receive less attention

than their prokaryotic counterparts (Grattepanche et al. 2014). This discrepancy is particularly evident in freshwater ecosystems, where eukaryotic communities are underexplored compared to marine environments. Understanding their diversity and functional roles in lakes, including how they interact within complex food webs and how their populations fluctuate over time, is essential. Temporal variations in eukaryotic communities, influenced by factors, such as seasonality, nutrient availability and environmental changes, are likely key to understanding lake ecosystem dynamics. A larger-scale, time-resolved exploration of these microorganisms is crucial for uncovering their contributions to trophic interactions and ecosystem functioning in freshwater systems.

Historically, molecular environmental studies relied on cloning and Sanger sequencing of the small subunit (SSU) ribosomal RNA gene (18S rDNA) (Sanger and Coulson 1975). This method generated reads long enough to allow relatively accurate phylogenetic interpretations (López-García et al. 2001; Moon-Van Der Staay et al. 2001; Massana et al. 2004) for species identification and differentiation (Hadziavdic et al. 2014; Bradley et al. 2016). However, contemporary approaches predominantly utilise much shorter reads generated by Illumina sequencing. These shorter reads are routinely produced in the millions and maintain high base-calling accuracy, resulting in reliable sequence data (de Vargas et al. 2015). This technological advancement has allowed for the sequencing of a significant portion of the species existing within an environment, including those that are extremely rare (Logares et al. 2014) and has yielded numerous important insights into microbial ecology and function. Typically, studies relating to eukaryotic diversity have targeted V4, V9 hypervariable regions or ITS regions, reflecting the limitations of Illumina sequencing, which can only sequence up to almost 600 bp (Stoeck et al. 2010; Mahé et al. 2015; Boenigk et al. 2018). The small size of the amplified fragments results in amplicons with weakened phylogenetic signal (Dunthorn et al. 2014), which can complicate the taxonomic assignment at finer taxonomic ranks, obscuring our understanding of species-level diversity and ecological interactions.

In response to the limitations of short reads, long-read sequencing technologies, such as Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (ONT), have emerged, enabling the generation of DNA sequences exceeding 30 kb (Tedersoo et al. 2021). Although these technologies initially exhibit higher error rates than Illumina sequencing, the use of circular consensus sequencing (CCS), which relies on repeated reading of the same molecule to correct errors, allows PacBio to reach accuracy levels comparable to Illumina (Travers et al. 2010; Jamy et al. 2020). This improvement has enabled the first applications of long-read sequencing to assess environmental eukaryotic diversity. Long-read technology has made it possible to sequence nearly complete 18S rDNA gene, extended fragments that include portions of the 28S gene and even regions spanning from the 18S to the 28S rDNA (including ITS1, 5.8S and ITS2 regions). This technology has been employed to investigate both global and group-specific eukaryotic diversity (Heeger et al. 2018; Orr et al. 2018; Tedersoo et al. 2018; Tedersoo and Anslan 2019; Jamy et al. 2020; Gaonkar and Campbell 2024). For example, Jamy et al. (2022) were able to obtain an evolutionary framework of environmental diversity and infer habitat-preference evolution across the eukaryotic tree. Similarly, recent long-read metabarcoding studies have shown that long-read OTUs provide higher taxonomic resolution,

enabling the discovery of previously unrecognised clades (Chwalińska et al. 2025). However, comprehensive comparative studies directly assessing short- versus long-read metabarcoding of the same environmental samples remain limited. Overgaard et al. (2024) further highlighted the lack of comparison between short-read markers and full-length rRNA operons. Such comparative analyses would allow us to assess the relative strengths and limitations of each sequencing technology, identify any significant discrepancies in the results obtained from these different methods and uncover potential biases inherent to each sequencing approach.

Beyond improving taxonomic resolution, long-read metabarcoding also opens new perspectives for ecological studies of microbial eukaryotes. By enabling more precise identification of taxa, longer amplicons may facilitate the detection of cryptic diversity and improve the characterisation of community composition and seasonal dynamics (Heeger et al. 2018; Jamy et al. 2020). Such improvements are particularly relevant in lake ecosystems, where microbial eukaryotic communities exhibit strong temporal fluctuations and play key roles in trophic interactions, including grazing and parasitism (e.g. Weisse (2014); Van den Wyngaert et al. (2022)). Higher taxonomic resolution may, therefore, help reveal ecological processes that remain difficult to detect with short-read data, such as succession patterns within closely-related taxa or potential host–parasite relationships within microbial food webs. In this study, we therefore employed both short-read and long-read metabarcoding approaches to characterise eukaryotic diversity and seasonal dynamics in a lake ecosystem throughout an entire annual cycle and to directly compare the results obtained from each method.

Materials and methods

Environmental samples and DNA extraction

The study was conducted in Lake Aydat (Massif Central, France, 45°39'50"N, 2°59'04"E). It is a dimictic and eutrophic lake with a maximum depth of 15 m, situated at an altitude of 825 m. From 19/11/2018 to 07/01/2020, water samples were collected in the euphotic zone of the lake at a permanent station located at the deepest zone of the water column using an integrated sampler IWS III (HYDRO-BIOS, Altenholz, Germany). Sampling was performed monthly, except from 27/05/2019 to 20/06/2019 when it was done twice a week. Fourteen litres of water were filtered through 150 µm pore-size filters, the microbial biomass was collected on 0.8 µm pore-size polycarbonate filters (Millipore, Burlington, MA) at a very low vacuum to prevent cell damage (< 15 kPa) and filters were stored at -80 °C until nucleic acids extraction. Two filters were produced and subsequently processed for each sample.

DNA extraction was performed at the end of the sampling period. Filters were covered with the extraction buffer (50 mM glucose, 10 mM EDTA, 25 mM Tris). Samples were homogenised in a bead beater for 3 cycles, 30 Hz 30 s (Bead Beater Retsch, Haan, Germany) with 0.1 g of 0.1 mm glass beads (Sigma-Aldrich, Saint-Louis, MO). Tubes were then centrifuged for 1 min at 600 g. EDTA pH 8 and a lysozyme solution (final concentrations 0.1 M and 1.5 mg ml-1, respectively) were added to the supernatant and incubated at 37 °C for 30 min.

Then sodium dodecyl sulphate (1.8%, final), proteinase K (100 µg ml/l, final) and RNaseA (50 µg ml/l, final) were added and samples were incubated at 37 °C for 60 min. A cetyltrimethylammonium bromide (CTAB) solution (final concentration 1% in a 0.7 M NaCl solution) was added and samples were incubated at 65 °C for 10 min. Nucleic acids were extracted with phenol-chloroform-isoamyl alcohol (25:24:1); the aqueous phase containing nucleic acids was kept and purified by adding chloroform-isoamyl alcohol (24:1). After isopropanol (0.6 vol) addition, the nucleic acids were precipitated at 20 °C for 12 h. After centrifugation, the DNA pellet was ethanol rinsed and re-suspended in 20 µl of water.

Short-read amplification, sequencing and taxonomic assignment

Amplification of the V4 region of eukaryotes was performed using the universal primer U515F (Ellis et al. 2013) and the eukaryotic primer Ek-NSR951 (Mangot et al. 2013) (Table 1), resulting in an expected amplicon length of ~ 380 bp. This primer pair was selected to provide broad eukaryotic coverage: in silico analysis using TestPrime 1.0 (Klindworth et al. 2013) showed that it does not amplify prokaryotes and efficiently targets most eukaryotic groups, although coverage is limited for a few lineages, such as Excavata, Discoba and Haptophyta, with 951R ensuring eukaryote specificity (SILVA r138.2 RefNR database, default parameters).

Table 1. List of primers used in this study.

Primer	Sequence (5'-3')	Reference
U515F	GTGYCAGCMGCCGCGGTA	Ellis et al. (2013)
Ek-NSR951	TTGGYRAATGCTTTCGC	Mangot et al. (2013)
3NDf	GGCAAGTCTGGTGCCAG	Cavalier-Smith et al. (2009)
21R	GACGAGGCATTTGGCTACCTT	Schwelm et al. (2016)
22R	CCATTCATGCRCGTCACWART	Schwelm et al. (2016)

Each Polymerase Chain Reaction (PCR) was performed in a total volume of 30 µl containing 6 µl of 5x Green GoTaq Flexi buffer, 2.4 µl of 25 mM $MgCl_2$, 0.18 µl of GoTaqG2 Flexi DNA Polymerase (Promega, Madison, WI, 5 U/µl), 0.6 µl of 10 mM each dNTP, 0.3 µl of 50 mg/ml BSA and 0.6 µl of each 10 µM primer. The amplification conditions consisted of initial denaturation at 95 °C for 5 min, followed by 33 cycles of 1 min at 95 °C, 1 min at 55 °C and 1 min 30 s at 72 °C, with a final elongation of 10 min at 72 °C. The amplicons were purified and concentrated using the QIAquick Gel Extraction Kit (Qiagen, Hilden, Germany). Samples were then sequenced on a MiSeq Illumina instrument (2× 250 bp) (Biofidal, Vaulx-en-Velin, France), where indexing PCR and subsequent demultiplexing were performed by the sequencing facility.

The PANAM2 (Phylogenetic Analysis of Next-generation AMplicons v.2 - https://github.com/panammeb/PANAM2) pipeline (Taib et al. 2013) was used to process the sequencing data. The clean-up procedures consisted of eliminating sequences with ambiguous 'N' bases, those below 200 bp and chimeras (Fig. 1A). Chimera detection was performed on a per-sample basis after paired-end read merging. The sequences were then assembled and clustered into Operational Taxonomic Units (OTUs) with a similarity threshold of 97%. Finally, representative sequences of each OTU were taxonomically assigned

according to a phylogenetic analysis using the PANAM2 curated reference database enriched for freshwater clades description (https://github.com/panammeb/Freshwater-database) (Fig. 1B), using a conservative lowest common ancestor (LCA) approach that reports the deepest supported taxonomic level (limited to the genus level in the present study). Taxonomic assignments were subsequently harmonised using the PR2 database (version 5.0.0; Guillou et al. (2013)). For this purpose, the deepest taxonomic rank supported by the PANAM2 assignment was mapped on to the corresponding rank in the fixed nine-level taxonomy implemented in PR2, allowing standardised comparisons across OTUs. OTUs assigned to metazoans were discarded for this study. Only OTUs representing more than 0.005% (Bokulich et al. 2013) of the total reads (corresponding to a minimum of 59 reads per OTU) were kept, resulting in 747 OTUs in the final short-read (SR) dataset (Fig. 1A, B). The number of reads remaining after each major step of the short-read bioinformatic pipeline is provided in Suppl. material 1: table S1.

Long-read amplification and sequencing

We used two sets of eukaryotic universal primers to amplify a region covering 18S, ITS1, 5.8S, ITS2 and 28S (~ 4500 bp) to maximise the diversity obtained: 3NDf/21R and 3NDf/22R (Cavalier-Smith et al. 2009; Schwelm et al. 2016; Jamy et al. 2020) (Table 1). *In silico* testing using SILVA TestProbe 3.0 (Quast et al. 2013) showed that primers 3NDf, 21R and 22R matched against 87.5%, 85.4% and 85.5% of all eukaryotic sequences, respectively (SILVA r138.2 RefNR database, default parameters). The main differences in taxonomic coverage between the two reverse primers were observed for four eukaryotic groups. Primer 21R amplified a higher proportion of Excavata (23.5%) and Discoba (96%) than primer 22R (0% and 36%, respectively), whereas primer 22R showed slightly higher coverage for Ciliophora (82% vs. 73%) and Rhizaria (92% vs. 89%).

For each sample, two PCRs were carried out (one for each set of primers) and the PCR products were subsequently pooled. PCR amplicons were generated using a barcoded forward primer (3NDf) and two alternative barcoded reverse primers (21R or 22R), allowing multiplexing of samples. Each PCR was performed in a total volume of 30 µl containing 6 µl of 5x SuperFi II Buffer, 0.5 µl of Platinium SuperFi II DNA polymerase (Invitrogen, Waltham, MA), 0.6 µl of 10 mM each dNTP, 0.6 µl of 50 mg/ml BSA and 0.6 µl of each 10 µM primer. The amplification conditions consisted of initial denaturation at 94 °C for 2 min, followed by 35 cycles of 30 s at 94 °C, 1 min at 55 °C and 6 min at 72 °C, with a final elongation of 20 min at 72 °C. Samples were then sequenced on a Pacbio Sequel II system (Gentyane, Clermont-Ferrand, France) and demultiplexing of the pooled amplicons was performed by the sequencing facility.

Long-read sequence curation, clustering and taxonomic assignment

Long-read preprocessing steps are summarised in Fig. 1A. Circular consensus sequences (CCS) were generated from raw reads using CCS version 6.3.0 from the SMRTtools version 11.0.0.146107 with a minimum number of three passes, a minimum predicted accuracy of 0.99 and a minimum quality score of 3 after truncation, with all other settings set to default. CCS, as well as raw

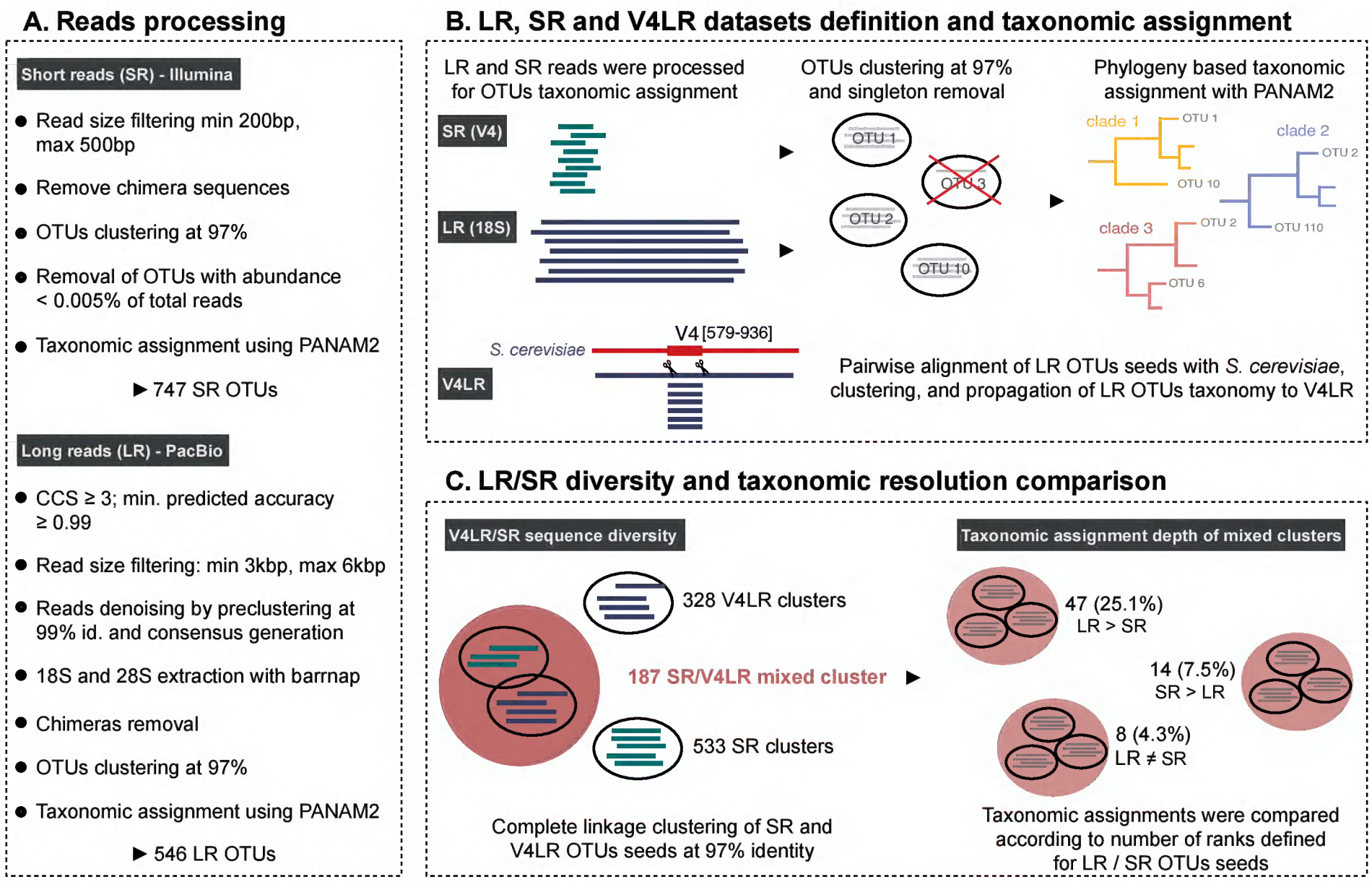
A. Reads processing
Short reads (SR) - Illumina
Read size filtering min 200bp, max 500bp
Remove chimera sequences
OTUs clustering at 97%
Removal of OTUs with abundance < 0.005% of total reads
Taxonomic assignment using PANAM2
► 747 SR OTUs
Long reads (LR) - PacBio
CCS ≥ 3; min. predicted accuracy ≥ 0.99
Read size filtering: min 3kbp, max 6kbp
Reads denoising by preclustering at 99% id. and consensus generation
18S and 28S extraction with barrnap
Chimeras removal
OTUs clustering at 97%
Taxonomic assignment using PANAM2
► 546 LR OTUs
B. LR, SR and V4LR datasets definition and taxonomic assignment
LR and SR reads were processed for OTUs taxonomic assignment
OTUs clustering at 97% and singleton removal
Phylogeny based taxonomic assignment with PANAM2
SR (V4)
LR (18S)
OTU 1
OTU 3
OTU 2
OTU 10
clade 1
clade 2
clade 3
OTU 1
OTU 2
OTU 10
OTU 2
OTU 110
OTU 6
S. cerevisiae
V4 [579-936]
V4LR
Pairwise alignment of LR OTUs seeds with S. cerevisiae, clustering, and propagation of LR OTUs taxonomy to V4LR
C. LR/SR diversity and taxonomic resolution comparison
V4LR/SR sequence diversity
328 V4LR clusters
187 SR/V4LR mixed cluster
533 SR clusters
Complete linkage clustering of SR and V4LR OTUs seeds at 97% identity
Taxonomic assignment depth of mixed clusters
47 (25.1%) LR > SR
14 (7.5%) SR > LR
8 (4.3%) LR ≠ SR
Taxonomic assignments were compared according to number of ranks defined for LR / SR OTUs seeds

Figure 1. Workflow for LR and SR processing and strategy for comparison of annotation depth. **A.** Long- and short-read processing leading to 747 and 546 SR and LR OTUs, respectively; **B.** Clustering threshold was set a 97% identity and taxonomic assignment was performed with PANAM2 (with a subsequent correction using PR2 to define fixed ranks). Pairwise alignments of LR OTUs seeds to the homologue reference of *S. cerevisiae* were performed resulting in V4LR sequences extraction and taxonomic assignment of LR has been applied to V4LR; **C.** OTU seeds from SR and V4LR datasets were clustered using a complete linkage clustering. A total of 187 mixed clusters were identified (clusters including both SR and V4LR OTU seeds). Taxonomic resolution was then compared in mixed clusters.

reads that did not result in CCS clean-up, was realised using a modified version of the PANAM2 preprocess step (https://github.com/meb-team/PANAM2_pacbio). In detail, PCR artefacts and low-quality sequences were removed using the filterAndTrim command in DADA2 (Callahan et al. 2016). Sequences with more than three expected errors and outside the 3000–6000 bp size range were removed. The sequences were then denoised according to a 99% identity preclustering with vsearch (version 2.15.2, Rognes et al. (2016)). Consensus sequences were generated for clusters with at least three sequences using mothur (version 1.48.0, Schloss et al. (2009)). Although the primer pair used amplifies nearly the entire ribosomal RNA operon, only the SSU (18S) region was used in this study. The concatenation of 18S and 28S data will be done in the future (Bronner et al., in prep). The 18S rDNA genes were, therefore, extracted with barrnap (version 0.9, https://github.com/tseemann/barrnap) and dereplicated before the removal of potential chimeras using the derep_fulllength and uchime_denovo commands from vsearch. Chimera detection was performed on a per-sample basis on the SSU regions extracted from barrnap. The sequences were then clustered into OTUs with a similarity threshold of 97% identity and taxonomically assigned using the same phylogenetic approach as the one used for short reads, resulting in 2,125 OTUs (Fig. 1B). The use of a

97% clustering threshold, although conservative for long-read data, was chosen to ensure consistency across datasets and comparability with other studies (Jamy et al. 2020, 2022). OTUs composed of a single sequence not emerging from a CCS, thus corresponding to a single read (singletons), were removed. OTUs classified as metazoans were also discarded, resulting in 546 OTUs in the final 18S long-read (LR) dataset (Figs 1A, B). The number of reads remaining after each major step of the long-read bioinformatic pipeline is provided in Suppl. material 1: table S2.

V4 extraction from long reads and clustering with short reads

To assess the impact of transitioning from long to short reads on the depth of taxonomic assignment in the same samples, we extracted the V4 region from the 2,125 LR seed sequences. The V4 regions were extracted by aligning the LR sequences on the *Saccharomyces cerevisiae* sequence AY251629.1 with the needleall function from the EMBOSS suite (Rice et al. 2000), then trimming the LR sequences fraction that aligned to the [579–936] interval of the *Saccharomyces cerevisiae* sequence, which correspond to the V4 region, using the Biostrings (version 2.66.0, Pagès et al. (2025)) R package (R version 4.2.2, R Development Core Team (2012)) (Fig. 1B). This process enabled us to obtain 2,107 V4 regions of average size 390 bp from 2,125 LR sequence (later referred as V4LR), resulting in 544 final OTUs (sequences were clustered as described for the short reads). The taxonomic assignments of LR have been retained and applied to V4LR (Fig. 1B).

To further compare the sequence diversity captured by SR and V4LR amplicons, respectively, both SR and V4LR OTUs seed sequences were clustered according to a complete linkage clustering with mothur at a 97% identity threshold. Clusters composition (SR only, V4LR only or mixed SR/V4LR) and taxonomic assignments were then compared (Fig. 1C).

Results

In total, we obtained 153,172 filtered CCS reads spanning ~ 4,500 bp of the rDNA operon, from the 18S to the 28S rDNA genes. A total of 20,851 18S sequences with an average size of 1,231.91 bp were extracted. After processing, sequences were clustered into Operational Taxonomic Units (OTUs) at 97% similarity, resulting in 2,125 OTUs. After removing singletons and OTUs classified as metazoans, 546 OTUs were kept in the final long-read (LR) dataset. The short-read (SR) dataset included 747 OTUs with an average sequence size of 380 bp.

SR/ V4LR comparison

To compare the extent of the genetic diversity captured by short and long sequencing strategies, we applied a complete linkage clustering (97% id) on pooled OTUs seeds of the short reads and the V4 regions extracted from the long reads (V4LR, whose taxonomic assignments had been retained from LR). A total of 1,048 clusters were obtained, of which 328 and 533 are composed only of V4LR and SR OTUs, respectively. Interestingly, the only-SR clusters are mostly composed of

OTUs with low abundance, compared to clusters that aggregate OTUs from both SR and LR (SR Wilcoxon test: p-value = 7.08e-14, Fig. 2). Amongst the 187 clusters containing OTUs from both SR and V4LR datasets, 118 (63.1%) show identical taxonomic assignments. Amongst these identical taxonomic assignments, 72% range from domain to subdivision only (1–4 taxonomic ranks), suggesting taxa whose taxonomic assignment is not differentially resolved by the sequencing method. Eight clusters (4.3% of mixed clusters) exhibit different taxonomic assignments. The discrepancies in these assignments are not associated with any particular taxonomic group. The 61 remaining clusters show varying degrees of taxonomic congruence especially due to the differences in taxonomy resolution. Indeed, 14 (7.5% of mixed clusters) clusters have a more detailed taxonomy in the SR dataset and 47 (25.1% of mixed clusters) in the LR dataset. Thus, more than a quarter of the clusters that shared SR/V4LR OTUs demonstrate a more precise taxonomic classification when using the full 18S sequence. When SR provides a more detailed classification, the taxonomic depth ranges from 1 to 4 ranks deeper (with an average increase of 1.8 ranks). For example, three clusters were assigned only to Rhizaria with LR, but could be further classified as Cercozoa with SR. In contrast, when LR offers more comprehensive classifications, the taxonomic depth extends from 1 to 7 ranks beyond SR (3.1 on average). The enhanced taxonomic resolution of LR sequencing is particularly evident in specific eukaryotic groups. Amongst Alveolates, LR consistently provides more granular classifications, especially for ciliates, where identification frequently reaches the family or genus level. Indeed, 65% of the clusters belonging to the ciliates are assigned up to the family and 35% up to the genus using LR (no taxonomic assignment beyond the class in SR). Similarly, Stramenopiles demonstrate significant improvement in taxonomic resolution, with OTUs classified only as Stramenopiles in SR being identified to specific diatom genera using LR.

Diversity and taxonomic resolution in SR and LR datasets

The overall eukaryotic diversity captured by both SR and LR datasets is relatively similar at the supergroup level, with a comparable number of OTUs assigned to each of them. However, notable differences emerge when examining the relative abundances of these supergroups. For instance, the Cryptista supergroup demonstrates a twofold higher relative abundance in the LR dataset compared to SR. Conversely, the Archaeplastida supergroup shows the opposite trend, with a higher relative abundance in the SR dataset.

Nine divisions are identified in both datasets. Tubulinea and Metamonada representing 0.3% (3 OTUs) and 0.03% (1 OTU) of total reads, respectively, are specific to LR, while the Discosea division is identified only in SR, representing 0.08% of total reads (1 OTU).

Differences in diversity become more apparent when considering the class level. Twenty-two classes are shared by both datasets; however, five are identified only in SR (Cryptophyceae, Centramoebia, Endomyxa-Phytomyxea, Colpodea, Pirsoniales) representing a total of six OTUs (0.26% of total reads) and 11 are specific to LR (Gyrista_X, Euglenida, Mediophyceae, Phyllopharyngea, Ascomycota, Armophorea, Coccidiomorphea, Zoopagomycota, Elardia, Prostomatea_1, Trichomonadea) representing a total of 58 OTUs (9.3% of total reads).

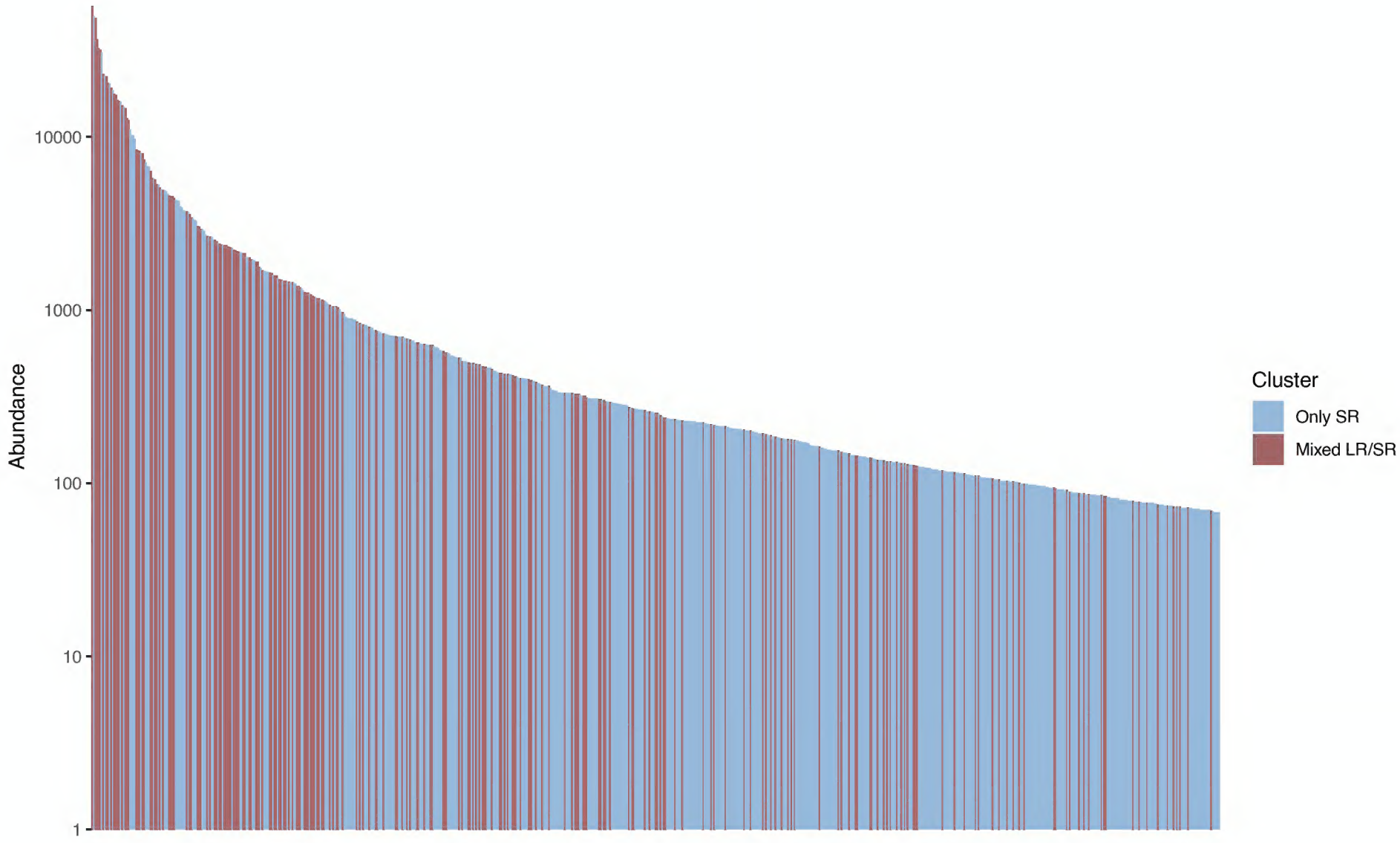

Figure 2. Abundances of OTUs in SR-only and mixed clusters. OTUs (x-axis) are coloured according to their inclusion in SR-only or mixed clusters, abundances of these OTUs (y-axis) are log-transformed (log_{10}). OTUs including in SR-only clusters show a significantly lower abundance (Wilcoxon test: p-value = 7.08e-14).

At a finer taxonomic level (i.e. genera), the differences are more striking; our analysis revealed 19 genera common to both SR and LR datasets (Fig. 3). The SR dataset exclusively identified 13 genera, with four belonging to the Chlorophyta division. These SR-specific genera encompassed 20 OTUs. Conversely, 35 genera are specific to the LR dataset (Fig. 3), comprising 87 OTUs and representing 13.7% of total reads. A significant proportion of these LR-specific genera are identified within the Fungi and Gyrista subdivisions. It is also at this rank that a shift in the classification depth is observed with 29.5% of OTUs from the LR dataset classified up to the genus level, compared with 11% in the SR dataset (Suppl. material 1: fig. S1).

When examining the most abundant subdivisions in both datasets, the classification is consistently improved in the LR dataset (Fig. 4). For example, none of the sequences belonging to the Apicomplexa is classified beyond this subdivision in SR, whereas 33.3% are classified down to the genus level in LR. Ciliates are also more accurately classified in the LR dataset. Indeed, more than half of their OTUs (59%) are not assigned beyond the phylum in SR (vs. 31% in LR) and 34% are assigned to a ciliate genus in LR (vs. 11% in SR) (Fig. 4). Actually, two genera of ciliates are identified in SR vs. 4 in LR (Fig. 3), including *Carchesium* as the second most represented (1.2% of total reads). Chytrids, which are abundant parasites in lake environments, show a significantly improved genus-level identification with long-read sequencing (36% vs. 8% in SR).

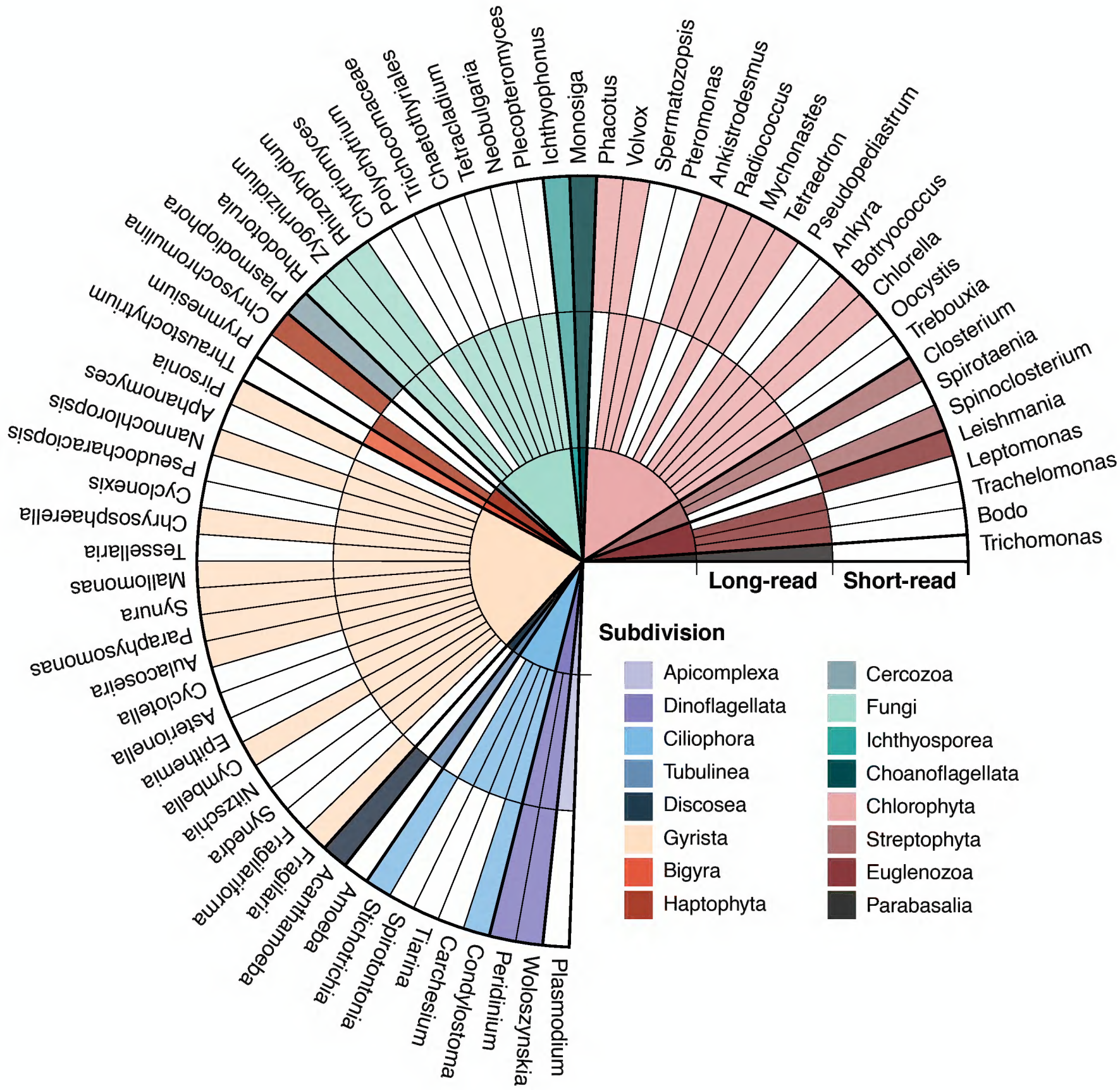

Figure 3. Genera identified in both datasets. From inner to outer circles: taxonomic subdivision (or division) of each identified genus, presence/absence in the LR dataset and presence/absence in the SR dataset. The colour hue corresponds to the taxonomic subdivision of the genus. In the outer circles, coloured segments indicate that the genus was detected in the corresponding dataset, whereas blank segments indicate its absence.

Amongst the Gyrista subdivision, diatom identification at the genus level is significantly enhanced in the LR dataset. While 92% of diatom OTUs are classified to genus level in LR, only 56% achieve this resolution in SR. The LR approach also reveals greater diatom diversity, identifying eight genera versus three in SR (Fig. 3). Notably, SR diatoms are predominantly assigned to *Fragilaria*, whereas LR uncovers substantial populations of *Cyclotella* (6.3% of total reads) and *Asterionella*, amongst others (Fig. 3).

The Dinoflagellata subdivision is the only one with a similar taxonomic depth between the two datasets (50% and 57% of OTUs assigned to the genus level in SR and LR, respectively) and the Cercozoa subdivision is the only one for which the taxonomic assignments are more precise in SR (3.3% of OTUs assigned to a genus, 0% in LR) (Fig. 4).

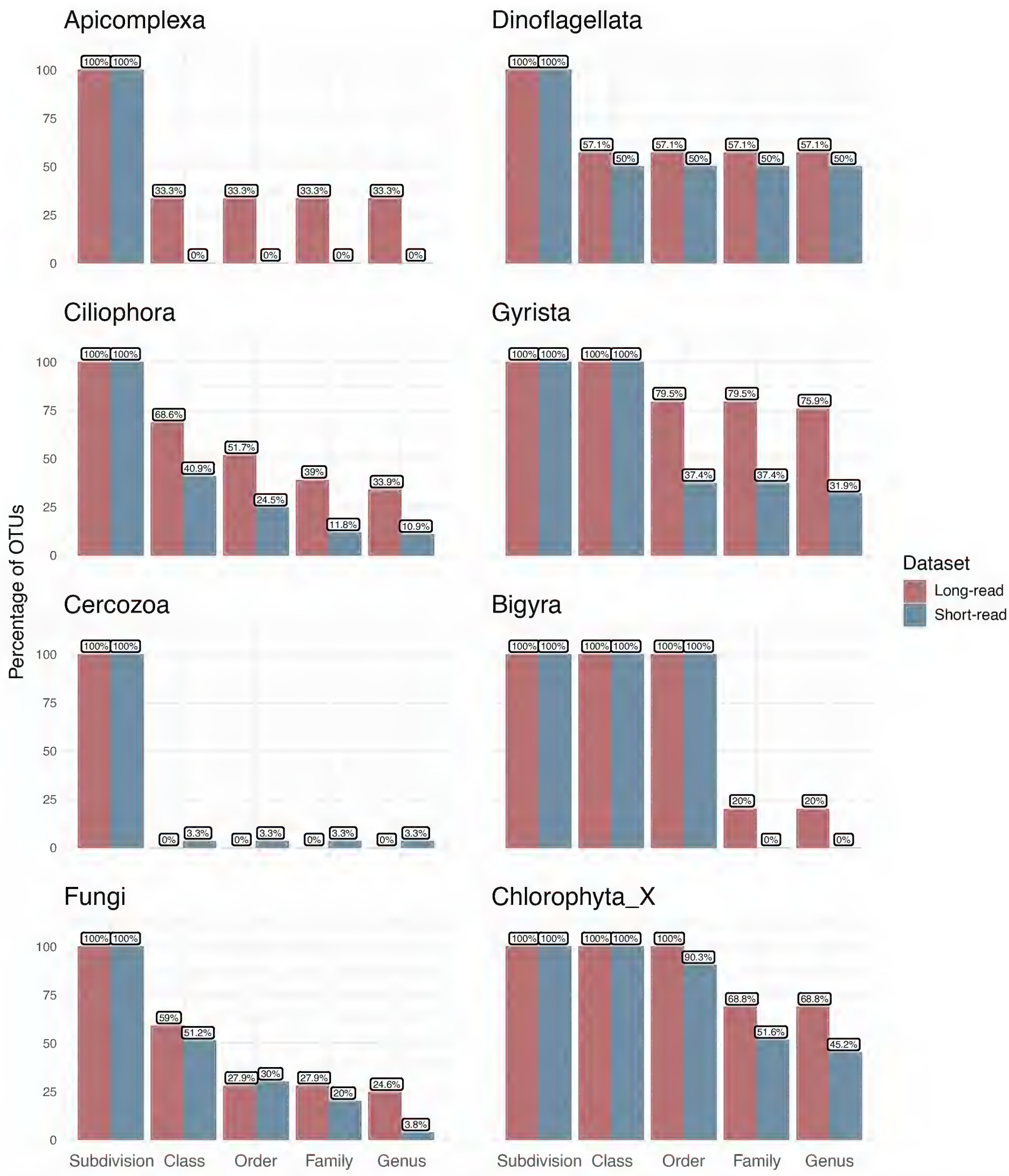

Figure 4. Percentage of OTUs assigned to each taxonomic level in both datasets within the eight most abundant subdivisions.

Dynamic of eukaryotic microbial diversity in Lake Aydat

Both datasets were used to analyse the dynamic of eukaryotic microbial diversity over a year in Lake Aydat (France). The most abundant subdivisions (defined as representing over 1% of total reads in both datasets) are Bigyra, Gyrista, Fungi, Dinoflagellata, Ciliophora, Chlorophyta and Cercozoa. With a higher taxonomic resolution, the LR dataset gives a more detailed dynamic

for Stramenopiles and Alveolata (Fig. 5). Relative abundance of Stramenopiles classes, such as Bacillariophyceae, Chrysophyceae and Mediophyceae, varied strongly over time, with marked fluctuations even between consecutive sampling dates (Fig. 5). Mediophyceae, for instance, were abundant from April to June, but almost completely absent from October to December. Within Alveolata, the SR dataset grouped most reads as “unclassified Ciliophora” (Suppl. material 1: fig. S2), whereas the LR dataset resolved finer temporal patterns for the ciliate classes Spirotrichea and Oligohymenophorea. Spirotrichea were, on average, more abundant over the year, but displayed

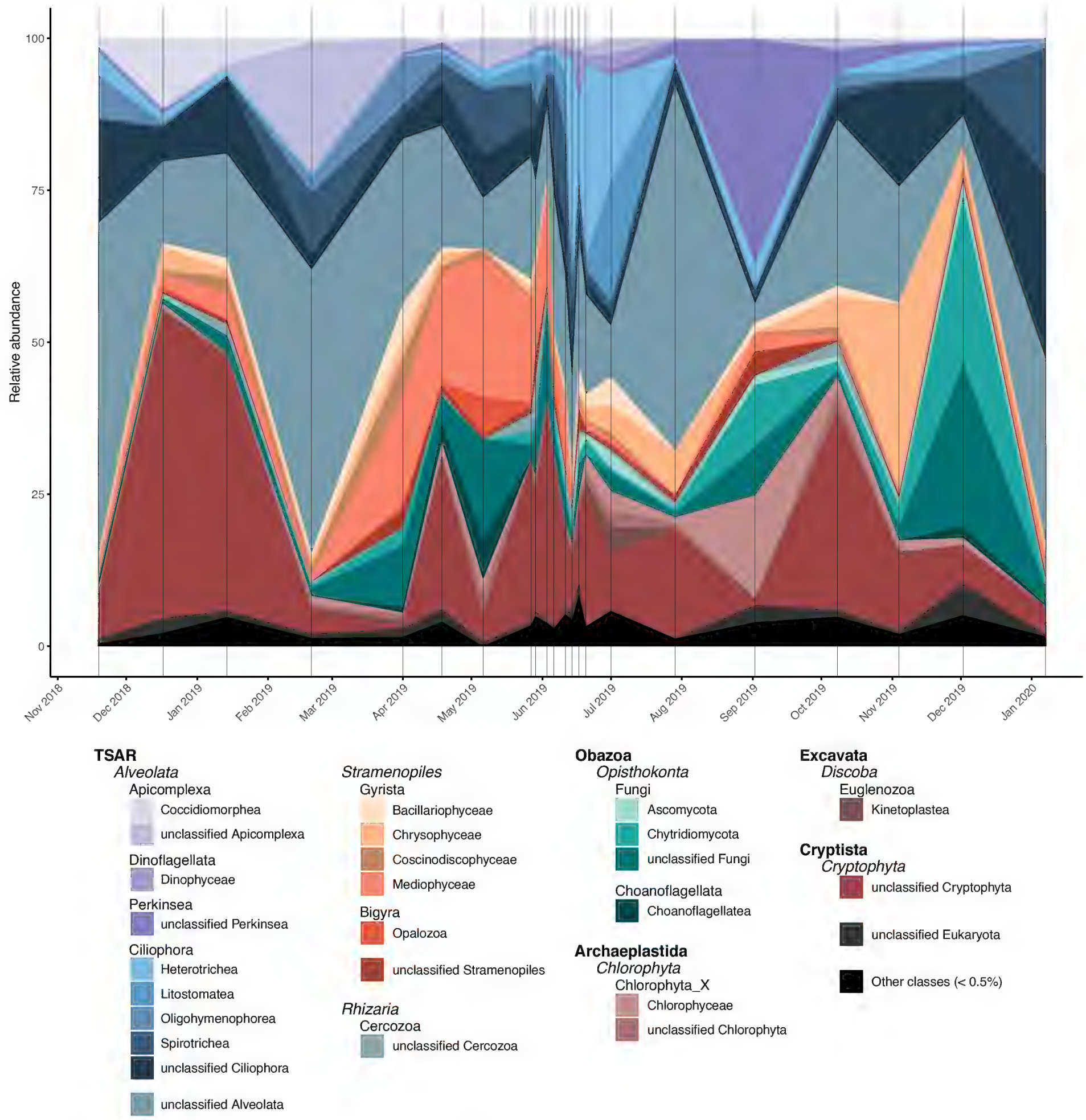

Figure 5. Temporal evolution of the relative abundances of the most abundant eukaryotic classes throughout the studied period using long reads. The black zone on the bottom represents the cumulative proportion of other classes with low abundances (< 0.5% of the total reads). Vertical bars represent sampling dates.

different dynamics from Oligohymenophorea (Fig. 2). The relative abundance of Oligohymenophorea peaked in November 2018 (7% of reads) and April 2019 (8%), while the remaining sampling dates showed an average of 1.2%. In contrast, Spirotrichea reached their highest relative abundances in June 2019 and January 2020. Fungal diversity is also globally better captured in the LR dataset; however, both datasets present the same temporal dynamics (Fig. 5, Suppl. material 1: fig. S2) with specifically an increase of their abundance during the colder months in 2019. On specific dates, major differences in terms of groups identified can be observed between the two datasets. For example, in September 2019, a peak of Perkinsea and Chlorophyceae are observed in LR, whereas in SR at the same date, more than 90% of the sequences are assigned to Litostomatea (ciliates) or unassigned beyond the eukaryotic domain (Fig. 5, Suppl. material 1: fig. S2). In March 2019, an important peak of Alveolata (not classified further) can be observed in LR, whereas a large part of the sequences seems to belong to Stramenopiles (not classified further) in SR (Fig. 5, Suppl. material 1: fig. S2).

Long-read sequencing allowed to highlight temporal variations at the genera level (Fig. 6). For example, the ciliate *Condylostoma* (class Heterotrichea) shows continuous presence, with a marked peak in June 2019. Diatom genera such as *Asterionella*, *Fragilariforma* (both Bacillariophyceae), *Aulacoseira* (Coscinodiscophyceae) and *Cyclotella* (Mediophyceae) are present throughout the year, but they exhibit distinct patterns of relative abundance. *Cyclotella* is favoured during the spring months (March to June), whereas the other three genera show low abundance following a peak in March. In July, *Fragilariforma* is the most abundant genus overall, while *Asterionella* reaches its highest abundance in January and April (Fig. 6). *Mallomonas* algae (Chrysophyceae) remain relatively stable, but exhibit a pronounced increase in November 2019. Some genera, such as the euglenozoan *Leptomonas* (Kinetoplastea), are present throughout the time series, but remain at lower relative abundances (Fig. 6). Finally, other genera display more seasonal or sporadic occurrences. For example, *Thraustochytrium* (Sagenista) and *Oocystis* (Trebouxiophyceae) are mostly observed in spring and summer. However, even more transient patterns are seen for the dinoflagellate *Peridinium* (Dinophyceae) almost only detected in June and July, and for *Plasmodium* (Coccidiomorphea) peaking between November 2018 and March 2019 (Fig. 6). *Zygorhizidium* (Chytridiomycota) remains almost undetected, except from March to May 2019 and from October to January 2020 (Fig. 6).

Discussion

Our study provides a year-long assessment of microbial eukaryotic diversity in a freshwater lake ecosystem (Lake Aydat, France), using both short-read (Illumina) and long-read (PacBio) metabarcoding approaches. A key outcome of our analysis is the improvement in taxonomic resolution achieved through long-read sequencing data. Over 25% of clusters containing SR and V4LR OTUs showed a deeper taxonomic assignment using LR data, with some cases showing up to seven additional hierarchical levels resolved. These findings corroborate previous studies highlighting the limited phylogenetic resolution of short amplicons (V4/V9 regions) which often result in ambiguous taxonomic

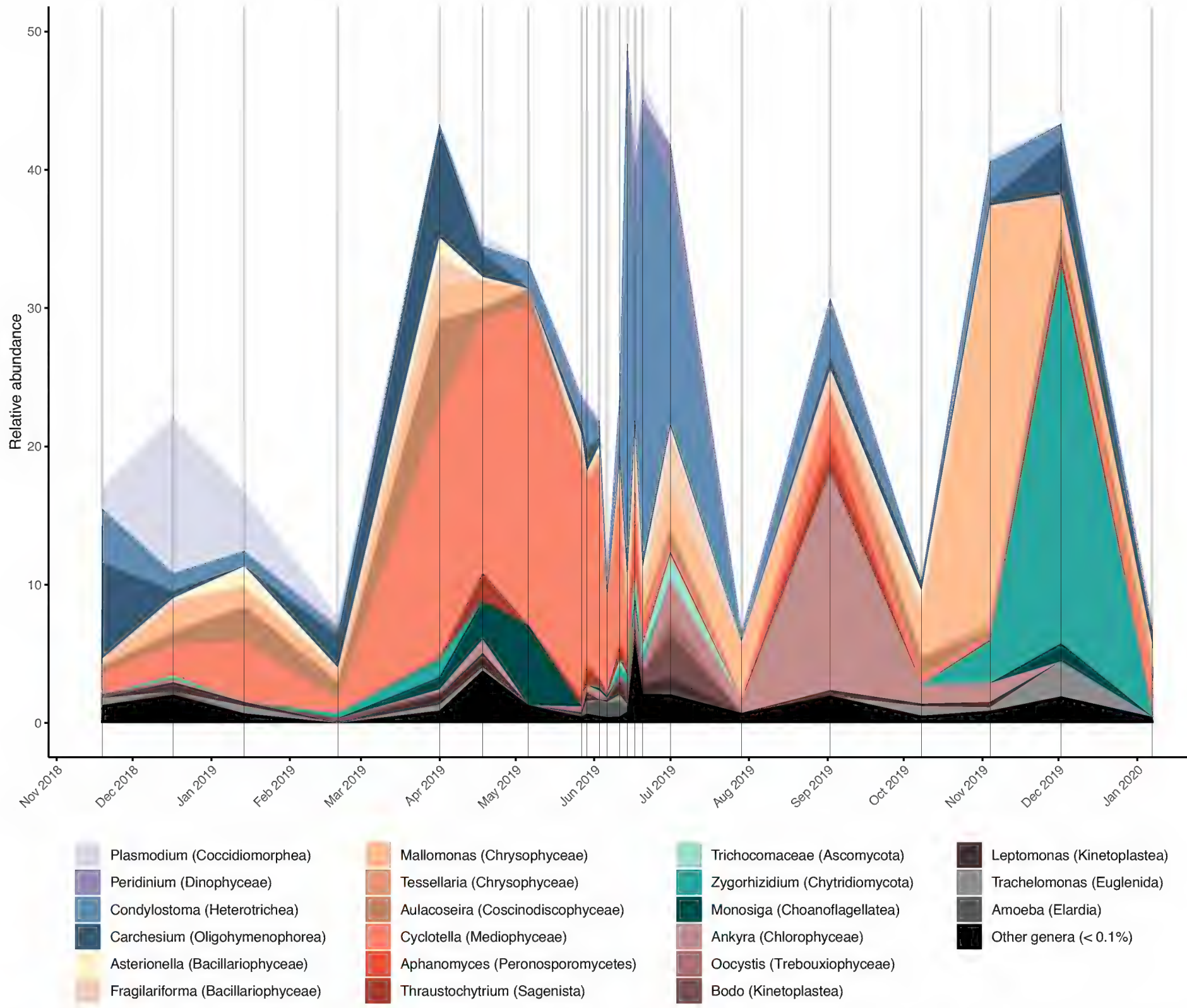

Figure 6. Temporal evolution of the relative abundances of the most abundant eukaryotic genera throughout the studied period using long reads. The black zone on the bottom represents the cumulative proportion of other genera with low abundances (< 0.1% of the total reads). Vertical bars represent sampling dates.

assignments at the species and genus levels (Taib et al. 2013) and may contribute to an overestimation of diversity due to the rapid evolution of hypervariable regions (Jamy et al. 2020; Guo et al. 2022). Despite these limitations, short-read approaches remain widely employed due to cost efficiency, well-established analytical pipelines and high sequencing depth, which enhances the detection of low-abundance taxa (likely explaining their predominance in SR-specific clusters in our dataset, Fig. 2). Taxonomic resolution improvements with LR sequencing were especially pronounced for groups, such as Ciliophora, Chytridiomycota, Bacillariophyta and Apicomplexa, consistent with observations reported by Jamy et al. (2020). This enhancement is primarily attributable to the inclusion of multiple variable regions (V1–V9) across the nearly full-length 18S rRNA gene, providing a greater number of informative sites for accurate taxonomic discrimination (Heeger et al. 2018; Tedersoo et al. 2021). Interestingly, in a limited number of cases, SR OTUs showed deeper taxonomic assignments than their LR counterparts. This pattern may reflect differences in reference database coverage, which strongly influence taxonomic

resolution and may favour well-represented short-read markers, such as the V4 region (Guillou et al. 2013; Jamy et al. 2020). In contrast, long-read sequences spanning multiple regions may lack close reference sequences across their full length, leading to more conservative assignments when using an LCA approach. Differential amplification efficiency between approaches may also contribute to these discrepancies.

The LR datasets captured a higher number of classes and genera, many of which were absent from the SR data. Notably, 35 genera were exclusively recovered by the LR approach, compared to only 13 unique genera identified solely by SR sequencing (Fig. 3). For example, LR sequencing uncovered a richer diversity of diatom genera, aligning well with microscopic observations (Chauvet et al. 2022), whereas SR data predominantly identified the genus *Fragilaria*. If followed by careful data processing, long-read sequencing proves highly effective in terms of both diversity detection and depth of taxonomic assignments (Krehenwinkel et al. 2019; Martijn et al. 2019; Jamy et al. 2020; Furneaux et al. 2021), as long as a good sequencing depth is achieved, enabling the least abundant organisms to be identified (Gaonkar and Campbell 2024). Taking this a step further, combining data from both the 18S and 28S rRNA genes can provide greater phylogenetic resolution within eukaryotic groups compared to using the 18S gene alone (Moreira et al. 2007; Zhao et al. 2012). For example, Marande et al. (2009), using Sanger sequencing of clone libraries, suggested that analysing both 18S and 28S genes allows for more reliable determination of the origins of environmental sequences, especially in rapidly evolving taxa. More recently, Jamy et al. (2022) leveraged sequences spanning 18S–28S rRNA to resolve phylogenetic relationships amongst environmental eukaryotes, providing a framework to link taxonomic diversity with habitat preferences, an approach that highlights the ecological insights possible with long-read data. It should also be noted that, while the long-read primer pair used in this study allows near-full-length amplification of the rRNA operon, long markers require careful handling of input DNA, as degraded or low-quality DNA can reduce amplification success. These considerations should be taken into account when designing long-read metabarcoding surveys for environmental monitoring. In addition, continually updated genomic databases enhance long-read sequencing by providing robust references that improve the accuracy of taxonomic assignments and support ecological interpretations.

The temporal patterns revealed by long-read sequencing provided a more detailed and nuanced understanding of seasonal dynamics within key eukaryotic groups in the lake ecosystem. Notably, the enhanced taxonomic resolution afforded by long-read data enabled us to detect contrasting seasonal dynamics of ciliates genera within the Oligohymenophorea, Spirotrichea and Heterotrichea classes. Ciliates are recognised as pivotal players in microbial food webs, yet their ecology remains underexplored due to challenges in accurate identification (Weisse 2014). For instance, we observed a succession between *Condylostoma* (Heterotrichea) and *Carchesium* (Oligohymenophorea) over the course of the year, despite both genera sharing similar dietary preferences (feeding on bacteria, small algae and other protists (Haraguchi et al. 2018)). Such temporal partitioning may reflect differences in life history strategies or environmental tolerances. The abundance and community composition of these ciliates has been found to be sensitive to fluctuations in nutrient

concentrations (particularly phosphorus and nitrogen) as well as other environmental variables, such as temperature and dissolved oxygen (Barouillet et al. 2022). This, together with accurate taxonomic assignment through long-read sequencing, could enhance their potential as bioindicators for monitoring ecosystem health and water quality.

Similarly, LR sequencing revealed temporal changes in the diversity of chytrid fungi providing evidence for seasonal turnover within genera such as *Zygorhizidium* (Fig. 6) and *Chytriomyces* (data not shown). These findings are ecologically significant, as chytrids are prominent parasites of phytoplankton. By infecting and lysing their hosts, they can influence the composition and succession of algal communities, sometimes terminating blooms or altering primary production patterns (Van den Wyngaert et al. 2022). The enhanced taxonomic resolution provided by long-read sequencing not only offers a more detailed view of the microbial eukaryotic community, but also facilitates the investigation of ecological interactions at finer taxonomic scales. For instance, this approach allows us to track the dynamics of the chytrid genus *Zygorhizidium* alongside various diatom genera, which are primary targets for chytrid parasitism. While laboratory studies have demonstrated that *Zygorhizidium affluens* can infect *Asterionella formosa* (Rad-Menéndez et al. 2018), such host-parasite relationships are challenging to study in natural environments. The use of long-read sequencing could help detect and monitor these interactions *in situ*. For example, our results indicate that the abundance of *Zygorhizidium* increases markedly between March and May, which may represent a response to the diatom bloom during this period (Fig. 6). Additionally, the pronounced peak of chytrids observed in December could be the result of a significant increase in *Mallomonas*, which subsequently led to their decline, as previously reported by Li et al. (2022). Finally, we also showed concomitant peaks in abundance of Perkinsea and Chlorophyceae (Fig. 5) that can confirm previous work highlighting that freshwater Perkinsea could play an important role as parasites of colonial Chlorophyceae (Jobard et al. 2020) while they are known primarily in marine environments as parasites of molluscs, crustaceans and dinoflagellates (Reñé et al. 2021).

Conclusions

This study demonstrates the potential of long-read sequencing for exploring the diversity and temporal dynamics of microbial eukaryotes in aquatic ecosystems. By directly comparing short-read and long-read metabarcoding approaches in Lake Aydat over an annual cycle, we show that long-read sequencing offers superior taxonomic resolution, enabling more precise identification of microbial eukaryotes. This enhanced resolution is particularly valuable for tracking seasonal shifts within key groups, such as ciliates and chytrid fungi, whose ecological roles are often obscured by the limitations of traditional and short-read sequencing methods. As sequencing costs continue to fall, the systematic integration of long-read sequencing into routine environmental monitoring is expected to resolve long-standing taxonomic ambiguities and enable the investigation of population dynamics at finer taxonomic resolution. Moreover, targeted cultivation, single-cell genomics and the growth of full-length, taxonomically validated reference genome collections will be crucial to fully exploit long-read datasets and to close remaining taxonomic gaps.

Acknowledgements

The authors are grateful to Gentyane sequencing platform (INRAE, Clermont-Ferrand, France) for the sequencing, as well as the Mésocentre Clermont Auvergne University and AuBi platform for providing help, computing and storage resources. We also thank Hélène Gardon for her help in the conceptualisation and data analysis and Arthur Monjot for his help with the figures. Authors acknowledge the financial support from the 'Fédération des Recherches en Environnement' and the Institut Universitaire de France (IUF).

References

Barouillet C, Vasselon V, Keck F, Millet L, Etienne D, Galop D, Rius D, Domaizon I (2022) Paleoreconstructions of ciliate communities reveal long-term ecological changes in temperate lakes. Scientific Reports 12: 7899. https://doi.org/10.1038/s41598-022-12041-7

Boenigk J, Wodniok S, Bock C, Beisser D, Hempel C, Grossmann L, Lange A, Jensen M (2018) Geographic distance and mountain ranges structure freshwater protist communities on a European scale. Metabarcoding and Metagenomics 2: e21519. https://doi.org/10.3897/mbmg.2.21519

Bokulich NA, Subramanian S, Faith JJ, Gevers D, Gordon JI, Knight R, Mills DA, Caporaso JG (2013) Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. Nature Methods 10: 57–59. https://doi.org/10.1038/nmeth.2276

Bradley IM, Pinto AJ, Guest JS (2016) Design and Evaluation of Illumina MiSeq-Compatible, 18S rRNA Gene-Specific Primers for Improved Characterization of Mixed Phototrophic Communities. Applied and Environmental Microbiology 82: 5878–5891. https://doi.org/10.1128/AEM.01630-16

Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP (2016) DADA2: High-resolution sample inference from Illumina amplicon data. Nature Methods 13: 581–583. https://doi.org/10.1038/nmeth.3869

Cavalier-Smith T, Lewis R, Chao EE, Oates B, Bass D (2009) *Helkesimastix marina* n. sp. (Cercozoa: Sainouroidea superfam. n.) a Gliding Zooflagellate of Novel Ultrastructure and Unusual Ciliary Behaviour. Protist 160: 452–479. https://doi.org/10.1016/j.protis.2009.03.003

Chauvet M, Debroas D, Moné A, Dubuffet A, Lepère C (2022) Temporal variations of Microsporidia diversity and discovery of new host–parasite interactions in a lake ecosystem. Environmental Microbiology 24(3): e15950. https://doi.org/10.1111/1462-2920.15950

Chwalińska M, Karlicki M, Romac S, Not F, Karnkowska A (2025) From short to long reads: enhanced protist diversity profiling via Nanopore metabarcoding. Metabarcoding and Metagenomics 9: e163750. https://doi.org/10.3897/mbmg.9.163750

Debroas D, Hugoni M, Domaizon I (2015) Evidence for an active rare biosphere within freshwater protists community. Molecular Ecology 24: 1236–1247. https://doi.org/10.1111/mec.13116

Dunthorn M, Otto J, Berger SA, Stamatakis A, Mahé F, Romac S, de Vargas C, Audic S, Biomarks Consortium, Stock A, Kauff F, Stoeck T (2014) Placing environmental next-generation sequencing amplicons from microbial eukaryotes into a phylogenetic context. Molecular Biology and Evolution 31: 993–1009. https://doi.org/10.1093/molbev/msu055

Ellis RJ, Bruce KD, Jenkins C, Stothard JR, Ajarova L, Mugisha L, Viney ME (2013) Comparison of the Distal Gut Microbiota from People and Animals in Africa. Ravel J (Ed.). PLoS ONE 8: e54783. https://doi.org/10.1371/journal.pone.0054783

Furneaux B, Bahram M, Rosling A, Yorou NS, Ryberg M (2021) Long- and short-read metabarcoding technologies reveal similar spatiotemporal structures in fungal communities. Molecular Ecology Resources 21: 1833–1849. https://doi.org/10.1111/1755-0998.13387

Gaonkar CC, Campbell L (2024) A full-length 18S ribosomal DNA metabarcoding approach for determining protist community diversity using Nanopore sequencing. Ecology and Evolution 14: e11232. https://doi.org/10.1002/ece3.11232

Grattepanche J-D, Santoferrara LF, McManus GB, Katz LA (2014) Diversity of diversity: conceptual and methodological differences in biodiversity estimates of eukaryotic microbes as compared to bacteria. Trends in Microbiology 22: 432–437. https://doi.org/10.1016/j.tim.2014.04.006

Guillou L, Bachar D, Audic S, Bass D, Berney C, Bittner L, Boutte C, Burgaud G, de Vargas C, Decelle J, Del Campo J, Dolan JR, Dunthorn M, Edvardsen B, Holzmann M, Kooistra WHCF, Lara E, Le Bescot N, Logares R, Mahé F, Massana R, Montresor M, Morard R, Not F, Pawlowski J, Probert I, Sauvadet A-L, Siano R, Stoeck T, Vaulot D, Zimmermann P, Christen R (2013) The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. Nucleic Acids Research 41: D597–604. https://doi.org/10.1093/nar/gks1160

Guo M, Yuan C, Tao L, Cai Y, Zhang W (2022) Life barcoded by DNA barcodes. Conservation Genetics Resources 14: 351–365. https://doi.org/10.1007/s12686-022-01291-2

Hadziavdic K, Lekang K, Lanzen A, Jonassen I, Thompson EM, Troedsson C (2014) Characterization of the 18S rRNA Gene for Designing Universal Eukaryote Specific Primers. PLoS ONE 9: e87624. https://doi.org/10.1371/journal.pone.0087624

Haraguchi L, Jakobsen HH, Lundholm N, Carstensen J (2018) Phytoplankton Community Dynamic: A Driver for Ciliate Trophic Strategies. Frontiers in Marine Science 5: 00272. https://doi.org/10.3389/fmars.2018.00272

Heeger F, Bourne EC, Baschien C, Yurkov A, Bunk B, Spröer C, Overmann J, Mazzoni CJ, Monaghan MT (2018) Long-read DNA metabarcoding of ribosomal RNA in the analysis of fungi from aquatic environments. Molecular Ecology Resources 18: 1500–1514. https://doi.org/10.1111/1755-0998.12937

Heger TJ, Giesbrecht IJW, Gustavsen J, del Campo J, Kellogg CTE, Hoffman KM, Lertzman K, Mohn WW, Keeling PJ (2018) High-throughput environmental sequencing reveals high diversity of litter and moss associated protist communities along a gradient of drainage and tree productivity. Environmental Microbiology 20: 1185–1203. https://doi.org/10.1111/1462-2920.14061

Jamy M, Biwer C, Vaulot D, Obiol A, Jing H, Peura S, Massana R, Burki F (2022) Global patterns and rates of habitat transitions across the eukaryotic tree of life. Nature Ecology & Evolution 6: 1458–1470. https://doi.org/10.1038/s41559-022-01838-4

Jamy M, Foster R, Barbera P, Czech L, Kozlov A, Stamatakis A, Bending G, Hilton S, Bass D, Burki F (2020) Long-read metabarcoding of the eukaryotic rDNA operon to phylogenetically and taxonomically resolve environmental diversity. Molecular Ecology Resources 20: 429–443. https://doi.org/10.1111/1755-0998.13117

Jobard M, Wawrzyniak I, Bronner G, Marie D, Vellet A, Sime-Ngando T, Debroas D, Lepère C (2020) Freshwater Perkinsea: diversity, ecology and genomic information. Beisner BE (Ed.). Journal of Plankton Research 42: 3–17. https://doi.org/10.1093/plankt/fbz068

Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, Glöckner FO (2013) Evaluation of general 16S ribosomal RNA gene PCR primers for classical and

next-generation sequencing-based diversity studies. Nucleic Acids Research 41: e1–e1. https://doi.org/10.1093/nar/gks808

Krehenwinkel H, Pomerantz A, Henderson JB, Kennedy SR, Lim JY, Swamy V, Shoobridge JD, Graham N, Patel NH, Gillespie RG, Prost S (2019) Nanopore sequencing of long ribosomal DNA amplicons enables portable and simple biodiversity assessments with high phylogenetic resolution across broad taxonomic scale. GigaScience 8: giz006. https://doi.org/10.1093/gigascience/giz006

Li L, Delgado-Viscogliosi P, Gerphagnon M, Viscogliosi E, Christaki U, Sime-Ngando T, Monchy S (2022) Taxonomic and functional dynamics during chytrid epidemics in an aquatic ecosystem. Molecular Ecology 31: 5618–5634. https://doi.org/10.1111/mec.16675

Logares R, Audic S, Bass D, Bittner L, Boutte C, Christen R, Claverie J-M, Decelle J, Dolan JR, Dunthorn M, Edvardsen B, Gobet A, Kooistra WHCF, Mahé F, Not F, Ogata H, Pawlowski J, Pernice MC, Romac S, Shalchian-Tabrizi K, Simon N, Stoeck T, Santini S, Siano R, Wincker P, Zingone A, Richards TA, de Vargas C, Massana R (2014) Patterns of rare and abundant marine microbial eukaryotes. Current Biology 24: 813–821. https://doi.org/10.1016/j.cub.2014.02.050

López-García P, Rodríguez-Valera F, Pedrós-Alió C, Moreira D (2001) Unexpected diversity of small eukaryotes in deep-sea Antarctic plankton. Nature 409: 603–607. https://doi.org/10.1038/35054537

Mahé F, Mayor J, Bunge J, Chi J, Siemensmeyer T, Stoeck T, Wahl B, Paprotka T, Filker S, Dunthorn M (2015) Comparing High-throughput Platforms for Sequencing the V4 Region of SSU-rDNA in Environmental Microbial Eukaryotic Diversity Surveys. The Journal of Eukaryotic Microbiology 62: 338–345. https://doi.org/10.1111/jeu.12187

Mangot J-F, Domaizon I, Taib N, Marouni N, Duffaud E, Bronner G, Debroas D (2013) Short-term dynamics of diversity patterns: evidence of continual reassembly within lacustrine small eukaryotes. Environmental Microbiology 15: 1745–1758. https://doi.org/10.1111/1462-2920.12065

Marande W, López-García P, Moreira D (2009) Eukaryotic diversity and phylogeny using small- and large-subunit ribosomal RNA genes from environmental samples. Environmental Microbiology 11: 3179–3188. https://doi.org/10.1111/j.1462-2920.2009.02023.x

Martijn J, Lind AE, Schön ME, Spiertz I, Juzokaite L, Bunikis I, Pettersson OV, Ettema TJG (2019) Confident phylogenetic identification of uncultured prokaryotes through long read amplicon sequencing of the 16S-ITS-23S rRNA operon. Environmental Microbiology 21: 2485–2498. https://doi.org/10.1111/1462-2920.14636

Massana R, Balagué V, Guillou L, Pedrós-Alió C (2004) Picoeukaryotic diversity in an oligotrophic coastal site studied by molecular and culturing approaches. FEMS Microbiology Ecology 50: 231–243. https://doi.org/10.1016/j.femsec.2004.07.001

Massana R, Gobet A, Audic S, Bass D, Bittner L, Boutte C, Chambouvet A, Christen R, Claverie J-M, Decelle J, Dolan JR, Dunthorn M, Edvardsen B, Forn I, Forster D, Guillou L, Jaillon O, Kooistra WHCF, Logares R, Mahé F, Not F, Ogata H, Pawlowski J, Pernice MC, Probert I, Romac S, Richards T, Santini S, Shalchian-Tabrizi K, Siano R, Simon N, Stoeck T, Vaulot D, Zingone A, de Vargas C (2015) Marine protist diversity in European coastal waters and sediments as revealed by high-throughput sequencing. Environmental Microbiology 17: 4035–4049. https://doi.org/10.1111/1462-2920.12955

Moon-van der Staay SY, De Wachter R, Vaulot D (2001) Oceanic 18S rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity. Nature 409: 607–610. https://doi.org/10.1038/35054541

Moreira D, von der Heyden S, Bass D, López-García P, Chao E, Cavalier-Smith T (2007) Global eukaryote phylogeny: Combined small- and large-subunit ribosomal DNA trees support monophyly of Rhizaria, Retaria and Excavata. Molecular Phylogenetics and Evolution 44: 255–266. https://doi.org/10.1016/j.ympev.2006.11.001

Orr RJS, Zhao S, Klaveness D, Yabuki A, Ikeda K, Watanabe MM, Shalchian-Tabrizi K (2018) Enigmatic Diphyllatea eukaryotes: culturing and targeted PacBio RS amplicon sequencing reveals a higher order taxonomic diversity and global distribution. BMC Evolutionary Biology 18: 115. https://doi.org/10.1186/s12862-018-1224-z

Overgaard CK, Jamy M, Radutoiu S, Burki F, Dueholm MKD (2024) Benchmarking long-read sequencing strategies for obtaining ASV -resolved rRNA operons from environmental microeukaryotes. Molecular Ecology Resources 24: e13991. https://doi.org/10.1111/1755-0998.13991

Pagès H, Aboyoun P, Gentleman R, DebRoy S (2025) Biostrings. Bioconductor. http://bioconductor.org/packages/Biostrings/ [January 5, 2026]

Pawlowski J, Audic S, Adl S, Bass D, Belbahri L, Berney C, Bowser SS, Cepicka I, Decelle J, Dunthorn M, Fiore-Donno AM, Gile GH, Holzmann M, Jahn R, Jirků M, Keeling PJ, Kostka M, Kudryavtsev A, Lara E, Lukeš J, Mann DG, Mitchell EAD, Nitsche F, Romeralo M, Saunders GW, Simpson AGB, Smirnov AV, Spouge JL, Stern RF, Stoeck T, Zimmermann J, Schindel D, Vargas C de (2012) CBOL Protist Working Group: Barcoding Eukaryotic Richness beyond the Animal, Plant, and Fungal Kingdoms. PLOS Biology 10: e1001419. https://doi.org/10.1371/journal.pbio.1001419

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Research 41: D590–596. https://doi.org/10.1093/nar/gks1219

R Development Core Team (2012) R: A language and environment for statistical computing. https://doi.org/10.32614/r.manuals

Rad-Menéndez C, Gerphagnon M, Garvetto A, Arce P, Badis Y, Sime-Ngando T, Gachon CMM (2018) Rediscovering *Zygorhizidium affluens* Canter: Molecular Taxonomy, Infectious Cycle, and Cryopreservation of a Chytrid Infecting the Bloom-Forming Diatom *Asterionella formosa*. Applied and Environmental Microbiology 84: e01826-18. https://doi.org/10.1128/AEM.01826-18

Reñé A, Timoneda N, Sampedro N, Alacid E, Gallisai R, Gordi J, Fernández-Valero AD, Pernice MC, Flo E, Garcés E (2021) Host preferences of coexisting Perkinsea parasitoids during coastal dinoflagellate blooms. Molecular Ecology 30: 2417–2433. https://doi.org/10.1111/mec.15895

Rice P, Longden I, Bleasby A (2000) EMBOSS: the European Molecular Biology Open Software Suite. Trends in genetics: TIG 16: 276–277. https://doi.org/10.1016/s0168-9525(00)02024-2

Rognes T, Flouri T, Nichols B, Quince C, Mahe F (2016) VSEARCH: a versatile open source tool for metagenomics. PeerJ 4: e2584. https://doi.org/10.7717/peerj.2584

Sanger F, Coulson AR (1975) A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. Journal of Molecular Biology 94: 441–448. https://doi.org/10.1016/0022-2836(75)90213-2

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Applied and Environmental Microbiology 75: 7537–7541. https://doi.org/10.1128/AEM.01541-09

Schwelm A, Berney C, Dixelius C, Bass D, Neuhauser S (2016) The Large Subunit rDNA Sequence of *Plasmodiophora brassicae* Does not Contain Intra-species Polymorphism. Protist 167: 544–554. https://doi.org/10.1016/j.protis.2016.08.008

Stoeck T, Bass D, Nebel M, Christen R, Jones MDM, Breiner H-W, Richards TA (2010) Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. Molecular Ecology 19(Suppl 1): 21–31. https://doi.org/10.1111/j.1365-294X.2009.04480.x

Sures B, Nachev M, Pahl M, Grabner D, Selbach C (2017) Parasites as drivers of key processes in aquatic ecosystems: Facts and future directions. Experimental Parasitology 180: 141–147. https://doi.org/10.1016/j.exppara.2017.03.011

Taib N, Mangot J-F, Domaizon I, Bronner G, Debroas D (2013) Phylogenetic Affiliation of SSU rRNA Genes Generated by Massively Parallel Sequencing: New Insights into the Freshwater Protist Diversity. PLoS ONE 8: e58950. https://doi.org/10.1371/journal.pone.0058950

Tedersoo L, Anslan S (2019) Towards PacBio-based pan-eukaryote metabarcoding using full-length ITS sequences. Environmental Microbiology Reports 11: 659–668. https://doi.org/10.1111/1758-2229.12776

Tedersoo L, Tooming-Klunderud A, Anslan S (2018) PacBio metabarcoding of Fungi and other eukaryotes: errors, biases and perspectives. The New Phytologist 217: 1370–1385. https://doi.org/10.1111/nph.14776

Tedersoo L, Albertsen M, Anslan S, Callahan B (2021) Perspectives and Benefits of High-Throughput Long-Read Sequencing in Microbial Ecology. Druzhinina IS (Ed.). Applied and Environmental Microbiology 87: e00626-21. https://doi.org/10.1128/AEM.00626-21

Travers KJ, Chin C-S, Rank DR, Eid JS, Turner SW (2010) A flexible and efficient template format for circular consensus sequencing and SNP detection. Nucleic Acids Research 38: e159–e159. https://doi.org/10.1093/nar/gkq543

Van den Wyngaert S, Ganzert L, Seto K, Rojas-Jimenez K, Agha R, Berger SA, Woodhouse J, Padisak J, Wurzbacher C, Kagami M, Grossart H-P (2022) Seasonality of parasitic and saprotrophic zoosporic fungi: linking sequence data to ecological traits. The ISME Journal 16(9): 2242–2254. https://doi.org/10.1038/s41396-022-01267-y

de Vargas C, Audic S, Henry N, Decelle J, Mahe F, Logares R, Lara E, Berney C, Le Bescot N, Probert I, Carmichael M, Poulain J, Romac S, Colin S, Aury J-M, Bittner L, Chaffron S, Dunthorn M, Engelen S, Flegontova O, Guidi L, Horak A, Jaillon O, Lima-Mendez G, Luke J, Malviya S, Morard R, Mulot M, Scalco E, Siano R, Vincent F, Zingone A, Dimier C, Picheral M, Searson S, Kandels-Lewis S, Tara Oceans Coordinators, Acinas SG, Bork P, Bowler C, Gorsky G, Grimsley N, Hingamp P, Iudicone D, Not F, Ogata H, Pesant S, Raes J, Sieracki ME, Speich S, Stemmann L, Sunagawa S, Weissenbach J, Wincker P, Karsenti E, Boss E, Follows M, Karp-Boss L, Krzic U, Reynaud EG, Sardet C, Sullivan MB, Velayoudon D (2015) Eukaryotic plankton diversity in the sunlit ocean. Science 348: 1261605–1261605. https://doi.org/10.1126/science.1261605

Weisse T (2014) Ciliates and the rare biosphere-community ecology and population dynamics. The Journal of Eukaryotic Microbiology 61: 419–433. https://doi.org/10.1111/jeu.12123

Worden AZ, Follows MJ, Giovannoni SJ, Wilken S, Zimmerman AE, Keeling PJ (2015) Rethinking the marine carbon cycle: Factoring in the multifarious lifestyles of microbes. Science 347: 1257594. https://doi.org/10.1126/science.1257594

Zhao S, Burki F, Bråte J, Keeling PJ, Klaveness D, Shalchian-Tabrizi K (2012) *Collodictyon*—an ancient lineage in the tree of eukaryotes. Molecular Biology and Evolution 29: 1557–1568. https://doi.org/10.1093/molbev/mss001

Supplementary material 1

Additional information

Authors: Marina Chauvet, Paul Terzian, Anne Moné, Didier Debroas, Gisèle Bronner, Cécile Lepère

Data type: pdf

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.10.184566.suppl1

Additional information

Conflict of interest

The authors have declared that no competing interests exist.

Ethical statement

No ethical statement was reported.

Artificial Intelligence (AI) use

The authors accept full responsibility for the content of the manuscript, including the disclosure of any use of AI.
No AI tools were used in the preparation of this manuscript.

Funding

This research was funded by the Fédération des Recherches en Environnement (FRE) and the Institut Universitaire de France (IUF).

Author contributions

Conceptualisation: CL, GB, DD, MC. Data curation: PT, MC. Formal analysis: MC. Funding acquisition: CL. Methodology: GB, DD, MC, CL. Writing – original draft: MC, CL. Writing – review and editing: MC, PT, CL, AM, DD, GB.

Author ORCIDs

M. Chauvet https://orcid.org/0000-0002-6380-6162
P. Terzian https://orcid.org/0000-0002-7092-3906
A. Moné https://orcid.org/0000-0002-8686-703X
D. Debroas https://orcid.org/0000-0002-9915-1268
G. Bronner https://orcid.org/0000-0002-4863-4518
C. Lepère https://orcid.org/0000-0003-4767-0477

Data availability

All of the data that support the findings of this study are available in the main text or Suppl. material 1. Short-read sequencing data were previously generated and published in Chauvet et al. (2022) and deposited under BioProject number PRJNA922126. The present study reuses a subset of these data for a different analytical purpose (from BioSample SAMN32643695 to SAMN32643737, excluding SAMN32643700, SAMN32643702, SAMN32643707, SAMN32643728 and SAMN32643734). Long-read sequencing data are available under BioProject number PRJEB108117.